IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Computational challenges and temporal dependence in Bayesian nonparametric models

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/1652936 since 2019-04-08T13:22:07Z

Published version:

DOI:10.1007/s10260-017-0397-8

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

04 May 2023

Noname manuscript No.
(will be inserted by the editor)

Computational challenges and temporal dependence in Bayesian nonparametric models

Raffaele Argiento · Matteo Ruggiero

Received: date / Accepted: date

Abstract ? provide an excellent review of several classes of Bayesian nonparametric models which have found widespread application in a variety of contexts, successfully highlighting their flexibility in comparison with parametric families. Particular attention in the paper is dedicated to modelling spatial dependence. Here we contribute by concisely discussing general computational challenges which arise with posterior inference with Bayesian nonparametric models and certain aspects of modelling temporal dependence.

Keywords Bayesian dependent model · Conjugacy · Computation · Dirichlet · Transition function

1 Computational challenges in Bayesian nonparametric models

One of the most successful strategies of the Bayesian nonparametric approach to statistical inference has arguably been semiparametric mixture modelling, which has proved to be extremely flexible and widely applicable. Semiparametric modelling assumes the observations are generated by parametric densities conditionally on the value of a set of parameters, which in turn are assigned a nonparametric distribution. More formally, we have the hierarchical representation

$$Y_i|\theta_i \overset{\text{ind}}{\sim} \varphi(y_i|\theta_i), \qquad \theta_i|G \overset{iid}{\sim} G, \qquad G \sim q(G^*, \zeta). \tag{1}$$

M. Ruggiero is supported by the European Research Council (ERC) through StG "N-BNP" 306406, and by the Italian Ministry of University and Research (PRIN 2015).

Raffaele Argiento and Matteo Ruggiero
University of Torino and Collegio Carlo Alberto
Corso Unione Sovietica 218/bis, 10134
Tel.: +39-011-670-6095/5758
E-mail: raffaele.argiento@unito.it; matteo.ruggiero@unito.it

Here $Y_1, \ldots, Y_n$ are the observations, $\theta_1, \ldots, \theta_n$ are a set of latent variables that parametrise the densities $\varphi(y_i|\theta_i)$, and G is a nonparametric distribution with prior q. The latter is in turn parametrised by a *baseline* distribution G^* on the parameter space $\Theta \subset \mathbb{R}^d$ and by a vector of reals ζ.

Upon observation of a dataset, posterior inference requires evaluating the conditional distribution of the parameters given the data. As is typically the case in absence of a fully conjugate model, i.e. such that the family of distributions assigned to the parameters is closed upon conditioning to the data, one needs to resort to Markov chain Monte Carlo methods to sample from the posterior. Early contributions dealing with this problem date back to **?**, **?** and **?**, and the large use of computer-aided inference has since boosted the investigation of new and efficient algorithms to deal with posterior analysis under a variety of modelling assumptions, generating a very lively literature. A brief introduction to such methods can focus on models as in (**??**) under the assumption that the mixing distribution G is almost surely a *discrete probability measure* with representation $G := \sum_{h \geq 1} w_h \delta_{\theta_h^*}$, where $\{w_h\}$ are random weights that sum up to one and $\{\theta_h^*\}$ are *iid* random points, taken independent of the weights, from the baseline distribution G^*. When the weights are obtained by normalising the increments of a time-changed subordinator, or more generally of a completely random measure (**?**), this specification coincides with the relevant class of random probability measures given by (homogeneous) *normalised random measures with independent increments* (**?**), which have recently been object of intense research and in turn include the celebrated Dirichlet process (**?**). See **?** for a recent review.

A broad classification of algorithms which enable to perform posterior inference under the above specifications divides them into *marginal* and *conditional* Gibbs sampling methods. Marginal Gibbs samplers are so called because they integrate out of (**??**) the random probability measure G. This entails sampling from

$$\mathcal{L}(\mathrm{d}\theta_1, \ldots, \mathrm{d}\theta_n|\mathbf{y}) \propto \prod_{i=1}^{n} \varphi(y_i|\theta_i)\mathcal{L}(\mathrm{d}\theta_1, \ldots, \mathrm{d}\theta_n) \tag{2}$$

where $\mathcal{L}(\mathrm{d}\theta_1, \ldots, \mathrm{d}\theta_n)$ is the prior marginal distribution of a sample from G and $\mathbf{y} = (y_1, \ldots, y_n)$ are the data. This marginal distribution can typically be characterised in terms of the predictive laws $\mathcal{L}(\theta_i|\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)$, which gives rise to Pólya urn schemes, in the case of the Dirichlet process (**?**), or generalisations thereof. Accordingly, Gibbs samplers with invariant distribution (**??**) are often called *generalised Pólya urns* Gibbs samplers.

Since G is discrete, the θ_i's will induce a partition $\rho = \{C_1, \ldots, C_k\}$ of $\{1, \ldots, n\}$ with $C_j = \{i : \theta_i = \theta_j^*\}$, where $j = 1, \ldots, k$ and $\theta_1^*, \ldots, \theta_k^*$'s are the distinct values in $\theta_1, \ldots, \theta_n$. Given that $\{x_h\}$ are *iid* and independent of $\{w_h\}$ in G, the law $\mathcal{L}(\theta_1, \ldots, \theta_n)$ is equivalent to

$$\mathcal{L}(C_1, \ldots, C_k, \theta_1^*, \ldots, \theta_k^*) = p(n_1, \ldots, n_k) \prod_{j=1}^{k} g^*(\theta_j^*),$$

where $n_i = \text{card}(C_i)$ and g^* is the density associated with G^*. The function p is called *exchangeable partition probability function* (EPPF), and represents the law of the random partition ρ. Marginalising over $\theta_1^*, \ldots, \theta_k^*$ yields back (**??**) which becomes

$$\mathcal{L}(C_1, \ldots, C_n|\mathbf{y}) \propto \prod_{j=1}^{k} m(y_{C_j}) p(n_1, \ldots, n_k) \tag{3}$$

where $m(y_{C_j}) := \int_\Theta \prod_{i \in C_j} \varphi(y_i|\theta) \mathrm{d}\theta$ is the marginal distribution of the data in group C_j. The computation now relies on the availability of efficient strategies for sampling from the EPPF, which are model-specific. Efficient solutions have been found based on the so called *Chinese restaurant process* (**?**) and its generalisations. Furthermore, expression (**??**) is the starting points for setting up inference under *product partition models* with regression on covariates, a class of models introduced in **?** and extended to the spatial setting by **?** (cf. also Section 5.3 of **?**).

The above, briefly described, marginal sampling methods are extremely useful since they allow to reduce an infinite-dimensional computation to a finite number of operations, entailed by integrating out the random probability measure. However, a downside is that inference is limited to point estimation of linear functionals of the population such as, e.g., predictive distributions, without allowing to quantify the associated uncertainty.

Alternative strategies retain the random probability measure G as part of the model, to be updated within the Gibbs sampling routine, and are therefore called *conditional* methods. Given the series representation of G, this strategies then shift the problem to that of simulating G, conditional on the data, with small or no approximation error. Truncation methods are the most intuitive option, and entail finding an appropriate N in $G_N := \sum_{h=1}^{N} w_h \delta_{x_h}$ which guarantees certain desired minimal requirements. Several ways to achieve these have been proposed, among others, in **??????**. These truncation methods are generally fairly easy to implement, but need to fix a priori, implicitly or explicitly, some notion of distance between the approximating and the target measure.

Other, very successful, stochastic truncation methods allow to perform exact sampling from the random probability measure and have proven to be reliable and relatively easy to implement. These include the *slice sampler* (**??**) and the *retrospective sampler* (**?**) together with their adaptations and generalisations. The slice sampler requires introducing appropriate latent [0,1]-valued variables u_i so that

$$\mathcal{L}(\mathrm{d}u_i, \mathrm{d}\theta_i|G) = \sum_{h=1}^{\infty} \mathbf{1}_{(0,w_h)}(\mathrm{d}u_i) \delta_{x_h}(\mathrm{d}\theta_i),$$

whereby integrating u out of the previous recovers $\mathcal{L}(\mathrm{d}\theta_i|G) = G(\mathrm{d}\theta_i)$. Define now

$$\mathcal{L}(\theta_i|u_i, G) = G_{u_i}(\mathrm{d}\theta_i) := \sum_{h \in A(u_i)} \delta_{x_h}(\mathrm{d}\theta_i)$$

where $A(u_i) := \{h : w_h > u\}$ is a finite set. From the latter it is clear that sampling G, conditional on $u_1, \ldots, u_n, \theta_1, \ldots, \theta_n$ and the data, entails updating only finitely-many of its components, namely the pairs (w_h, x_h) for $h \in \cup_{i=1}^{n} A(u_i)$.

The retrospective sampler instead is based on the idea of exchanging the intuitive order of simulation for sampling from G. This would lead to sampling the infinite sequences $\{w_h\}, \{x_h\}$, then draw v_i uniformly distributed in $(0, 1)$ and set $\theta_i = x_l$ if $\sum_{j=1}^{l-1} w_l < v_i < \sum_{j=1}^{l-1} w_l$, The retrospective sampler instead first samples v_i and then draws as many w_h, x_h as are needed to meet the above inequalities.

Gibbs sampling procedures described so far are very appealing strategies but still computationally intensive methods. This makes the use of mixtures such as (**??**) infeasible when dealing with large datasets, or when the computational resources are limited. Recently, variational Bayes methods have been proposed as an alternative (**?**). Acting essentially as optimisation algorithms, under these methods the posterior distribution of G is approximated by a distribution $\tilde{q}$, called *variational distribution*, of a finite dimensional process. The goal is then to adjust the parameters of $\tilde{q}$ in order to minimise the Kullback–Leibler divergence between $\tilde{q}$ and the posterior. Robustness of variational Bayes methods is currently one of the open problems in the Bayesian nonparametric literature, as it is known they can underestimate the model uncertainty.

2 Temporal dependence in Bayesian nonparametric models

An important line of research in Bayesian nonparametrics on the so called dependent processes has developed from the ideas introduced in **?**, where collections of dependent random probability measures $\{G_z, z \in \mathcal{Z}\}$ are considered, and G_z encodes the current state of the problem in correspondence of the covariate value z. Cf. **?**, Section 3.2.1. Computational methods for dependent models are very often problem-specific extensions of those summarised in Section **??**. Providing a general overview of these computational strategies would be a difficult task far beyond the scope and possibilities of this discussion. Since Section 5 of **?** presents some applications of dependent models for spatial data, we choose here to briefly discuss some issues related to models with temporal dependence, with particular emphasis on the role of conjugacy.

A common setting for Bayesian inference with temporal dependence is that of partial exchangeability, whereby the available data are of the form $y_{t_i,1}, \ldots, y_{t_i,n_i}$, where the indices t_i are discrete data-collection times, $n_i \geq 1$ for all i, and the data $y_{t_i,j}$ are such that, as j varies,

$$y_{t_i,j} \mid G_{t_i} \sim^{iid} G_{t_i}.$$

Hence the data are exchangeable across the t_i-sections, but not overall. From a temporal modelling perspective, one ideally wants the correlation between

pairs of random measures G_t and G_s increase as the indices t and s get closer, and decay to zero as t and s grow farther apart.

A non exhaustive list of contributions along this line of investigation which are based on Dirichlet mixture models includes, among others, **???????**. Other contributions have explored models which go beyond the structure of the Dirichlet process or closely related constructions, aiming at modelling, for example: marginal measures of the dependent process of geometric stick-breaking type (**?**), of Pitman–Yor type (**?**), of GEM type (**?**), or of gamma type (**?**); evolving binary matrices for relational network structures (**?**), or for dynamic feature allocation (**?**); monotonic functional time series (**?**); emission distributions for hidden Markov models (**??**).

Here we are interested in highlighting two roles conjugacy can play in these approaches to inference. One is with the aim of constructing stationary temporal models with a built-in simulation scheme available, as done in **???**. The kernel of the idea is to consider joint distributions

$$\mathcal{L}(\mathrm{d}\theta_1, \ldots, \mathrm{d}\theta_n, \mathrm{d}G) = \mathcal{L}(\mathrm{d}G) \prod_{i=1}^{n} \mathcal{L}(\mathrm{d}\theta_i \mid G)$$

where q is the nonparametric prior on G, and to construct transition functions through latent variables by writing

$$P(G, \mathrm{d}H) = \int \mathcal{L}(\mathrm{d}H \mid \theta_1, \ldots, \theta_n) \prod_{i=1}^{n} \mathcal{L}(\mathrm{d}\theta_i \mid G) \tag{4}$$

where $\mathcal{L}(\mathrm{d}H \mid \theta_1, \ldots, \theta_n)$ is the posterior of H given $\theta_1, \ldots, \theta_n$. For example, if $p := G(A) \in [0, 1]$ for some fixed set A, the law of $G(A)$ is a beta distribution and $\mathcal{L}(\mathrm{d}\theta_i \mid G(A))$ is Bernoulli with parameter p, then the above reduces to a beta-binomial transition

$$P(p, \mathrm{d}p') = \mathrm{beta}(\mathrm{d}p' \mid a + \theta, b + n - \theta)\mathrm{Binom}(\theta \mid n, p)$$

where (a, b) are prior beta hyperparameters. Note that this is in fact the transition function of the marginal state of a two-dimensional Gibbs sampler on the augmented space of (p, θ), which is stationary with respect to a beta. In a nonparametric framework, if $\mathcal{L}(\mathrm{d}G) = \Pi(\mathrm{d}G \mid \alpha)$ for some finite parameter measure α, and Π is conjugate in the sense that $G \sim \Pi$ and $X \mid G \sim G$ jointly imply $\Pi \mid X \sim \Pi(\cdot \mid f(\alpha, X))$ for some function f of the data and the prior parameter, then (**??**) yields

$$P(G, \mathrm{d}H) = \int \Pi(\mathrm{d}H \mid f(\alpha, \theta_1, \ldots, \theta_n)) \prod_{i=1}^{n} G(\mathrm{d}\theta_i). \tag{5}$$

Here Π can be shown to be the reversible measure of the process, so this strategy allows to construct stationary nonparametric processes. **?** discuss along these lines the Bayesian interpretation of the dynamics of two families of continuous-time Dirichlet and gamma dependent models for Bayesian

nonparametrics, the latter used for example in **?**. See also **?**. Their transition functions can be obtained by randomising n in (**??**) and by introducing appropriate coefficients which make $P(G, \mathrm{d}H)$ satisfy the Chapman–Kolmogorov conditions in continuous time. For example, for these two families one has

$$P_t(G, \mathrm{d}H) = \sum_{n\geq 0} P(N_t = n) \int \Pi(\mathrm{d}H \mid f(\alpha, \theta_1, \ldots, \theta_n)) \prod_{i=1}^{n} G(\mathrm{d}\theta_i), \qquad (6)$$

where N_t is an appropriate $\mathbb{Z}_+$-valued continuous-time process which determines the size of the latent sample $(\theta_1, \ldots, \theta_n)$. This approach has been followed explicitly in **?**. The resulting transitions are therefore infinite mixtures. Simulation of these transition functions can in principle be done by resorting to one of the methods outlined in the previous Section, e.g. by using a slice sampler twice on the mixture (**??**) and on the infinite-dimensional random measure which is the state of process, as done for example in **?**. Model-specific hurdles however may make this unfeasible, e.g. **?** develop an exact simulation scheme for (**??**) in the finite and infinite-dimensional Dirichlet cases, which deals efficiently with the non trivial expression for $P(N_t = n)$.

Alternatively, conjugacy can be deliberately sought in order to reduce the overall Monte Carlo error and predictive uncertainty within a broader computation. **?** for example extend classical posterior characterisations for Dirichlet and gamma random measures to the two above-mentioned families of dependent processes, conditional on discretely collected data. In particular, sufficient conditions are identified for these models (cf. **?**) that allow to write (**??**), conditional on $y_1, \ldots, y_m$ collected possibly at different times, as

$$P(G, \mathrm{d}H \mid y_1, \ldots, y_m) = \sum_{i=0}^{m} w_i(t) \Pi(\mathrm{d}H \mid f(\alpha, y_1, \ldots, y_i)).$$

This reduces (**??**), upon conditioning to the observed data, to a finite mixture of distributions in the same conjugate family. Note that the mixture components only consider $y_1, \ldots, y_i$ and not the entire sample. The $w_i(t)$'s are appropriate time-dependent weights which regulate how the posterior mass is reassigned at different times to the mixture components.

References

Arbel, J. and Prünster, I. (2017). A moment-matching Ferguson & Klass algorithm. *Stat. Comput.* **27**, 3–17.

Argiento, R., Bianchini, I. and Guglielmi, A. (2016a). A blocked Gibbs sampler for NGG-mixture models via a priori truncation. *Stat. Comput.* **26**, 641–661.

Argiento, R., Bianchini, I. and Guglielmi, A. (2016b). Posterior sampling from ε-approximation of normalized completely random measure mixtures. *Electron. J. Stat.* **10**, 3516–3547.

Barrios, E., Lijoi, A., Nieto-Barajas, L.E. and Prünster, I. (2013). Modeling with normalized random measure mixture models. *Stat. Sci.* **28**, 313–334.

Blackwell, D. and MacQueen, J.B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1**, 353–355.

Blei, D.M. and Jordan, M.I. (2006). Variational inference for Dirichlet process mixtures. *Bayesian anal.* **1**, 121-143.

Canale, A. and Ruggiero, M. (2016). Bayesian nonparametric forecasting of monotonic functional time series. *Electron. J. Stat.* **10**, 3265–3286.

Caron, F., Davy, M., Doucet, A., Duflos, E. and Vanheeghe, P. (2008). Bayesian inference for linear dynamic models with Dirichlet process mixtures. *IEEE Trans. Sig. Proc.* **56**, 71–84.

Caron, F., Davy. M. and Doucet, A. (2007) Generalized Polya urn for time-varying Dirichlet process mixtures. *Proc. 23rd Conf. on Uncertainty in Artificial Intelligence*, Vancouver.

Caron, F., Neiswanger, W., Wood, F., Doucet, A. and Davy, M. (2017). Generalized Pólya urn for time-varying Pitman-Yor processes. *J. Mach. Learn. Res.*, in press.

Caron, F. and Teh, Y.W.(2012). Bayesian nonparametric models for ranked data. *Neural Information Processing Systems* (NIPS 2012), Lake Tahoe, USA, 2012.

Dunson, D.B. (2006). Bayesian dynamic modelling of latent trait distributions. *Biostatistics* **7**, 551–568.

Durante, D. and Dunson, D. (2014). Nonparametric Bayes dynamic modelling of relational data. *Biometrika* **101**, 883–898.

Escobar, M.D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90**, 577–588

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.

Griffin, J.E. and Steel, M.F.J. (2010). Stick-breaking autoregressive processes. *J. Econometrics* **162**, 383–396.

Gutierrez, L., Mena, R.H. and Ruggiero, M. (2016). On GEM diffusive mixtures. In *JSM Proceedings, Section on Nonparametric Statistics*. Alexandria, VA: American Statistical Association.

Ishwaran, H. and James, L. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96**, 161–173.

Jenkins, P.A. and Spanò, D. (2017). Exact simulation of the Wright-Fisher diffusion. *Ann. Appl. Probab.*, in press.

Kingman, J.F.C. (1967). Completely random measures. *Pacific J. Math.* **21**, 59–78.

Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (N. L. Hjort, C. C. Holmes, P. Müller and S. G. Walker, eds.) 80–136. Cambridge Univ. Press, Cambridge.

Lijoi, A., Ruggiero, M. and Spanò, D. (2016). On the transition function of some time-dependent Dirichlet and gamma processes. In *JSM Proceedings, Section on Nonparametric Statistics*. Alexandria, VA: American Statistical Association.

Linderman, S.W., Johnson, M.J., Wilson, M.W. and Chen, Z.. A nonparametric Bayesian approach to uncovering rat hippocampal population codes during spatial navigation. *J. Neurosci. Meth.* **263**, 36–47.

Kalli, M. and Griffin, J. E. and Walker, S. G. (2011). Slice sampling mixture models. *Stat. Comput.* **21**, 93-105.

MacEachern, S.N. and Müller, P. (1998). Estimating mixture of Dirichlet process models. *J. Comp. Graph. Stat.* **7**, 223–238.

MacEachern, S.N. (1999). Dependent nonparametric Processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*. American Statist. Assoc., Alexandria, VA.

Mena, R.H. and Ruggiero, M. (2016). Dynamic density estimation with diffusive Dirichlet mixtures. *Bernoulli* **22**, 901–926.

Mena, R.H. and Walker, S.G. (2009). On a construction of Markov models in continuous time. *Metron* **67**, 303–323.

Mena, R.H., Ruggiero, M. and Walker, S.G. (2011). Geometric stick-breaking processes for continuous-time Bayesian nonparametric modelling. *J. Statist. Plann. Inf.* **141**, 3217–3230.

Muliere, P. and Tardella, L. (1998). Approximating distributions of random functionals of Ferguson-Dirichlet priors. *Canadian J. Statist.* **26**, 283–297.

Müller, P., Quintana, F. A. and Rosner, G. L. (2011) A product partition model with regression on covariates. *J. Comp. Graph. Stat.* **20**, 260–278.

Müller, P., Quintana, F.A. and Page, G. (2017). Nonparametric Bayesian inference in applications. *Stat. Methods Appl.* to be completed.

Neal, R.M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *J. Comp. Graph. Stat.* **9**, 249–265.

Page, G. and Quintana, F. A. (2016). Spatial Product Partition Models. *Bayes. Anal.* **11**, 265–298.

Papaspiliopoulos, O. and Ruggiero, M. (2014). Optimal filtering and the dual process. *Bernoulli* **20**, 1999–2019.

Papaspiliopoulos, O. and Roberts, G.O. (2008). Retrospective mcmc for dirichlet process hierarchical models. *Biometrika* **95**, 169–186.

Papaspiliopoulos, O., Ruggiero, M. and Spanò, D. (2016). Conjugacy properties of time-evolving Dirichlet and gamma random measures. *Electr. J. Statist.* **10**, 3452–3489.

Perrone, V., Jenkins, P.A., Spanò, D. and Teh, Y.W. (2017). Poisson random fields for dynamic feature models. Preprint available at *arXiv:1611.07460*.

Pitman, J. (2006). Combinatorial Stochastic Processes, *in École d'Été de Probabilités de Saint-Flour XXXII-2002*, Berlin: Springer-Verlag.

Pitt, M.K. and Walker, S.G. (2005). Constructing stationary time series models using auxiliary variables with applications. *J. Amer. Statist. Assoc.* **100**, 554–564.

Regazzini, E., Lijoi, A. and Prünster, I. (2003). Distributional results for means of random measures with independent increments. *Ann. Statist.* **31**, 560–585.

Rodriguez, A. and Ter Horst, E. (2008). Bayesian dynamic density estimation. *Bayes. Anal.* **3**, 339–366.

Ruggiero, M. and Walker, S.G. (2009). Bayesian nonparametric construction of the Fleming–Viot process with fertility selection. *Statist. Sinica*, **19**, 707–720.

Taddy, M. A. and Kottas, A. (2009). Markov switching Dirichlet process mixture regression. *Bayes. Anal.* **4**, 793–816.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Commun. Stat. Simul. Comput.* **36**, 45–54.

Walker, S.G., Hatjispyros S.J. and Nicoleris, T. (2007). A Fleming–Viot process and Bayesian nonparametrics. *Ann. Appl. Probab.* **17**, 67–80.

Yau, C., Papaspiliopoulos, O., Roberts, G.O. and Holmes, C. (2011). Bayesian non-parametric hidden Markov models with applications in genomics. *J. Roy. Statist. Soc. Ser. B* **73**, 37–57.